Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- La entropía de una variable aleatoria es una medida de la incertidumbre de la variable aleatoria:
  - es una medida de la cantidad de información requerida en promedio para describir la variable aleatoria.
- La entropía relativa es una medida de la distancia entre dos distribuciones de probabilidad.
- La entropía relativa D(p||q) es una medida de la ineficiencia de suponer que la distribución es q cuando la distribución verdadera es p.

Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- En términos de entropía, la entropía relativa o divergencia de Kullback la podemos entender así:
  - Si conociéramos la verdadera distribución $p(x)$ de una variable aleatoria, podríamos construir un código con una longitud de descripción promedio $H(p(x))$.
  - Sin embargo, si usáramos el código para una distribución $q(x)$ (que difiere de $p(x)$), necesitaríamos $H(p(x)) + D(p(x)||q(x))$ bits en promedio para describir la variable aleatoria.

Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- Así definimos la entropía relativa o distancia Kullback-Leibler entre dos funciones de probabilidad p(x) y q(x) se define como:
  - D (p || q) = ∑ p(x) log [p(x)/q(x)], donde el sumatorio se hace en los x є X.
- O lo que es lo mismo D (p || q) = Ep log [p(x)/q(x)] .
- Aquí usamos el convenio de que 0 log 0/0 = 0 y el convenio (basado en argumentos de continuidad) que 0 log 0/q = 0 y p log p/0=∞.
- Por tanto, si hay algún símbolo x ∈ X tal que p(x)> 0 y q(x) = 0, entonces D (p || q) = ∞.

Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- La distancia Kullback-Leibler entre dos funciones de probabilidad p(x) y q(x) es siempre es no negativa y es cero si y solo si p(x)=q(x).
- Sin embargo, no es una distancia propiamente dicha entre distribuciones ya que no es simétrica y además no satisface la desigualdad del triángulo:
  - la distancia satisface la desigualdad triangular: la longitud de un lado de un triángulo es menor que la suma de las longitudes de los otros dos lados.
- Aunque, no obstante, a menudo es muy útil pensar en la entropía relativa como una "distancia" entre distribuciones.

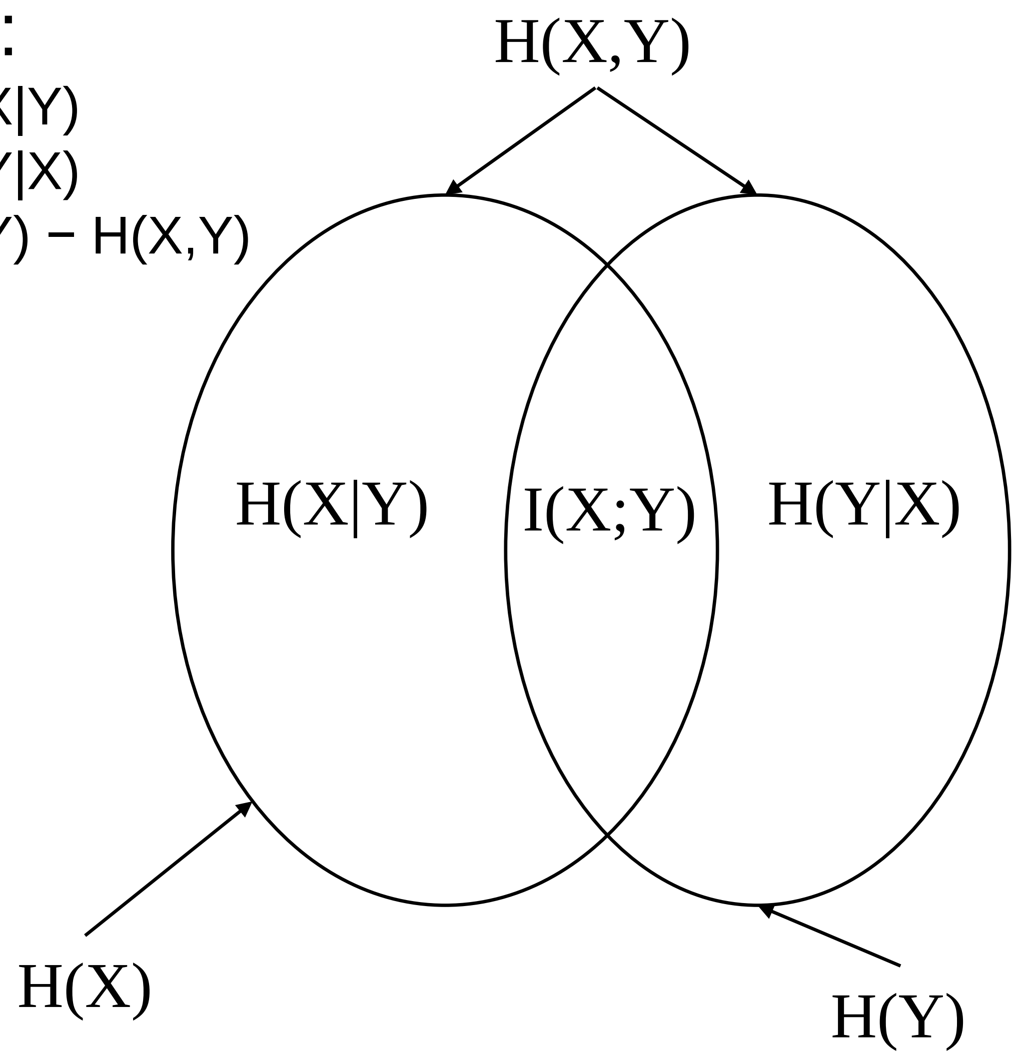
Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- La información mutua se pude definir en términos de la distancia Kullback-Leibler.
- Considere dos variables aleatorias X e Y con una probabilidad conjunta p(x,y) y probabilidad marginal p(x) y p(y). La información mutua I(X;Y) (lo que veníamos llamando hasta ahora MI (X;Y)) es la entropía relativa entre la distribución conjunta y la distribución del producto p(x) p(y):
  - $I(X;Y) = \sum \sum p(x,y) \log[p(x,y)/p(x)p(y)]$ donde el doble sumatorio se hace en los $x \in X$, $y \in Y$.
- Es decir $I(X;Y) = D(p(x,y)||p(x)p(y)) = E_{p(x,y)} \log[p(X,Y)/p(X)p(Y)]$.

Información mutua y sus medidas de dependencia: distancia Kullback–Leibler

- Ya sabemos que:
  - $I(X; Y) = H(X) - H(X|Y)$
  - $I(X; Y) = H(Y) - H(Y|X)$
  - $I(X; Y) = H(X) + H(Y) - H(X,Y)$
  - $I(X; Y) = I(Y; X)$
  - $I(X; X) = H(X)$.

La relación entre H (X), H(Y), H(X,Y), H(X|Y), H(Y|X), e I(X;Y) se expresa en este diagrama de Venn. Observe que la información mutua I(X;Y) corresponde a la intersección de la información en X con la información en Y.



BIBLIOGRAFÍA:

- T. M. Cover, and J. A. Thomas; Elements of Information Theory. Ed. J. Wiley and Sons, 1991.
- Information Theory, Inference and Learning Algorithms. David J.C. MacKay. (http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html\#book).
- Theoretical Neuroscience Computational and Mathematical Modeling of Neural Systems. Peter Dayan and LF Abbott. MIT 2001: Chapter 4 - Information Theory.

BIBLIOGRAFÍA:

- Fundamentals of Computational Neuroscience, Thomas P. Trappenberg. Oxford University Press. First edition, 2002: Chapter 5.
- Fundamentals of Computational Neuroscience, Thomas P. Trappenberg. Oxford University Press. Second edition 2010: Appendix D: Basic information theory.
- Spikes: Exploring the Neuronal Code. F. Rieke and D. Warland and R. de Ruyter van Steveninck and W. Bialek. A Bradford Book. MIT Press. Cambridge, Massachusetts, London, England, 1997: Chapter 3 - Quatifying information transmission

BIBLIOGRAFÍA:

- Fundamentos de estadística. Daniel Peña Sánchez Ribera. Alianza Editorial, 2001 o 2008.
- Data analysis Statistical and computational methods for scientists and engineers Recurso electrónico 4th.ed. Autor Brandt, Siegmund.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer Second Edition February 2009.

BIBLIOGRAFÍA:

- Analysis and study on text representation to improve the accuracy of the normalized compression distance. Ana Granados Fontecha 2012 (Doctoral Thesis).
- A. N. Kolmogorov (1968) Three approaches to the quantitative definition of information , International Journal of Computer Mathematics, 2:1-4, 157-168, DOI: 10.1080/00207166808803030
- Ming Li; Paul M.B. Vitányi (2009). An Introduction to Kolmogorov Complexity and Its Applications. Springer. pp. 105–106. ISBN 978-0-387-49820-1.

BIBLIOGRAFÍA:

- C.H. Bennett, P. Gács, M. Li, P.M.B. Vitányi, and W. Zurek. Information Distance, IEEE Transactions on Information Theory, 44:4(1998), 1407–1423.
- R. Cilibrasi and P. Vitanyi. Clustering by Compression. IEEE Transactions on Information Theory, 51(4):1523–1545, 2005.